DOCUMENT RESUME

ED 068 513 TM 001 851

TITLE	National Assessment Measuring American Education.
INSTITUTION	Education Commission of the States, Denver, Colo.
PUB DATE	72
NOTE	36p.
JOURNAL CIT	Compact; v6 n1 Feb 1972
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Academic Achievement; Educational Objectives; Evaluation Criteria; *Measurement Instruments; *National Competency Tests; Periodicals; *Program Effectiveness; *Student Evaluation; Testing
IDENTIFIERS	*National Assessment of Educational Progress

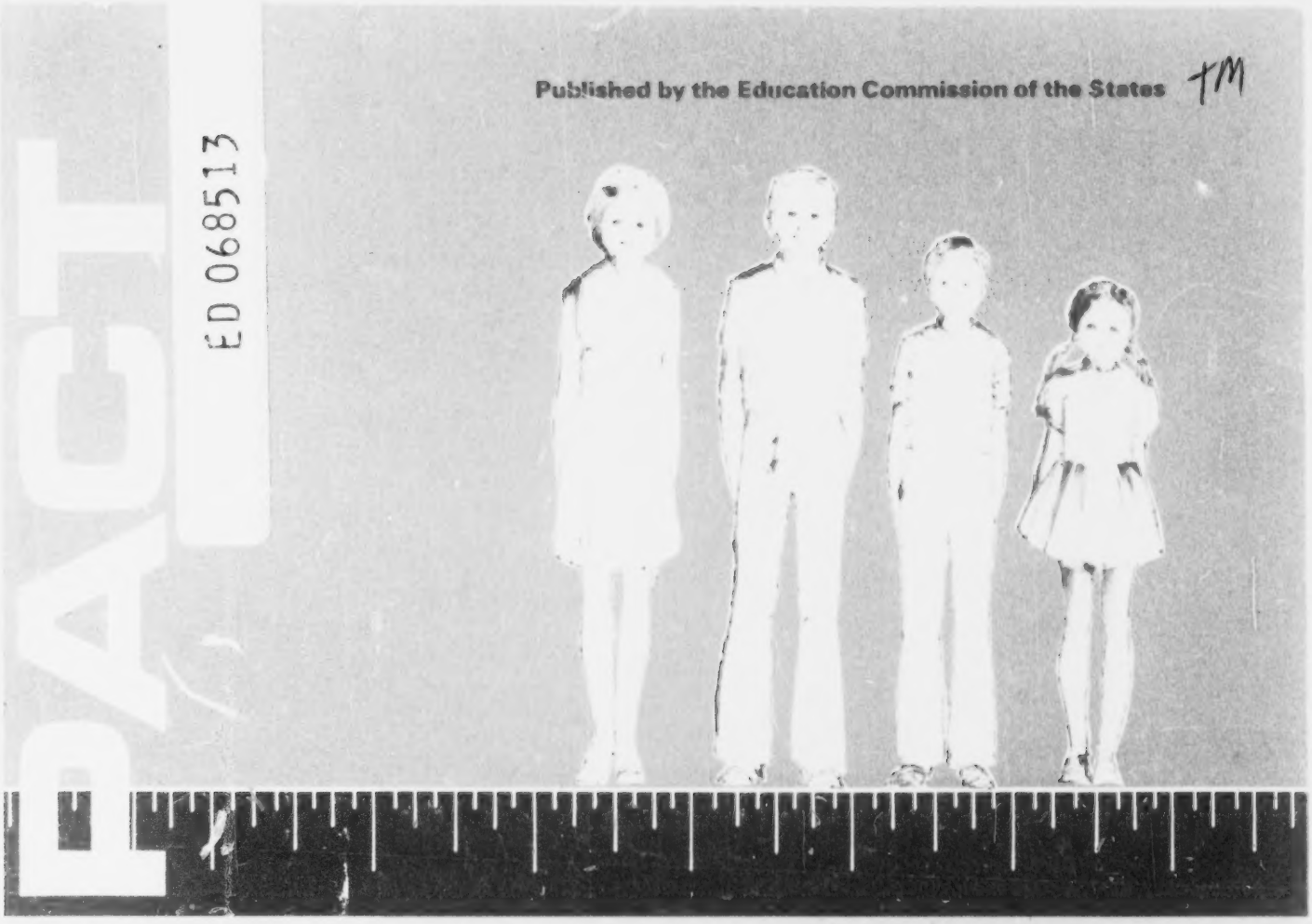
ABSTRACT

A selection of articles on National Assessment of Educational Progress, which measures American education, is presented with varying points of view. The introductory editorial reflects on the progress of National Assessment in measuring learning and in achieving its goal of sound rational decisions about the future of American education. The following articles deal with reasons for evaluating education, some values of assessment, industry as a consumer, the significance of National Assessment as compared to ordinary standardized tests, the first results, technical giants of the project, a layman's view, the possibility of changing American education, making the data work, trends in State assessment, and program support. Comments and quotes by individuals and publications about the program are also included. (LH)

ED 068513

Published by the Education Commission of the States

TM



NAPACT

NATIONAL ASSESSMENT

MEASURING AMERICAN EDUCATION

CG 008 514

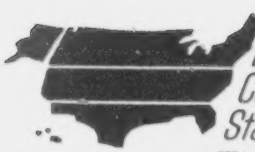
PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Education Commission of the States
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

The cover design and inside artwork for this issue of COMPACT were done by Mary Lou Egan, ECS art editor. She is a graduate of the University of Denver and holds her degree in advertising art.

COMPACT is published bimonthly by the Education Commission of the States to present a balanced viewpoint on the important educational issues of our times. Articles reflect the opinions of the authors and do not necessarily represent the position of the Commission. The annual subscription rate is $6.00. Single issues are available at $1.00 per copy. A reduced price is available for multiple copies, if mailed to the same address, at a rate of $.75 per copy for 10 to 49 issues, and $.50 per copy for 50 or more issues. On quantities of more than 150 copies, a special unit cost may be negotiated. Checks should be made payable to the Education Commission of the States and sent to ECS, 300 Lincoln Tower, 1860 Lincoln Street, Denver, Colorado 80203.

The Education Commission of the States, a nonprofit organization representing 41 states and two territories, is designed to provide a partnership between state education and political leaders for the advancement of education. The executive director is Wendell H. Pierce and Clifford L. Dochterman is director of communications.

Helen K. Masterson, Editor
Jeanine Bays, Associate Editor
Mary Lou Egan, Art Editor



Education Commission of the States
300 Lincoln Tower · 1860 Lincoln Street · Denver, Colorado 80203 (303) 893-5200

COMPACT

Copyright © The Education Commission of the States 1972. All rights reserved.

Contents

Francis Keppel—Marks on the Wall 2
Ralph W. Tyler—Why Evaluate Education? 3
George B. Brain—Some Values of Assessment 5
John K. Wolfe—Industry—An Unnoticed Consumer 7
Carmen J. Finley—Not Just Another Standardized Test 9
J. Stanley Ahmann—The First Results 13
Clay Allison—Technical Giants of National Assessment 18
Zoe von Ende—Layman's View 23
Frank B. Womer—Marjorie M. Mastic—Can National Assessment Change American Education? 26
George Johnson—Making the Data Work 29
Educational Testing Service—Trends in State Assessment 31
ECS—Support of Program 32
Comments and Quotes 24
Tables: Steps of National Assessment 17

EDITORIAL

Marks on the Wall

Watching the growth of National Assessment today reminds me of a family custom. Every so often, as the child grew, a mark was put on the kitchen door to record his height—and to show progress. As a nation, we are now asking for marks to measure the learning—and the progress, if any—of the coming generation. National Assessment is beginning to show its marks on the kitchen door, and the neighbors are taking notice.

Many of the fears expressed in the formative years that assessment inevitably would result in invidious comparisons have proven unfounded. In fact, earlier accusations that the ultimate goal of a nationally administered assessment was a plot to "take over" local or state assessments have now changed into requests from those very groups for help and advice. There can be little doubt that National Assessment of Educational Progress is becoming a credible and creditable measure of the progress of American education. And it is beginning to show that it can supplement and enhance state and local efforts.

The initial progress of National Assessment was typically American. It was started in the private sector and credit should go to the Carnegie Corporation of New York and the Ford Foundation for providing generous grants at the outset, and to the hundreds of professional and lay people, acting in their personal capacities, who carefully built the foundation on which National Assessment is now developing as a public enterprise.

However, as National Assessment moves into adolescence, let us not expect too many immediate "results." There will still be some stages of trial and error. If, as many of us hope, National Assessment is to become a major social indicator, there will have to be many periods for adjustment and improvement between now and then. The fact that the preliminary reports from the first data are already giving us information which is helpful to states, to the schools and to society-at-large shows its potential. But these first results are only the first benchmarks against which progress can be measured. We cannot and should not be expected to show progress, or lack of progress, until the second, and perhaps the third, reporting cycles have been completed. As in the case of growing children, the parents have to learn patience.

The goal of National Assessment is to provide for sounder and more rational decisions about the future of American education. All the data now gathered may be considered merely the first dot on a graph designed to provide American education with the necessary charts to achieve important social and educational goals.

Francis Keppel
Chairman of the Board
General Learning Corporation
U.S. Commissioner of Education (1962-65)

RALPH W. TYLER

Why Evaluate Education?

Dr. Tyler currently is senior consultant with Science Research Associates, Inc. in Chicago; director emeritus of the Center for Advanced Study in the Behavioral Sciences, Stanford, Palo Alto, Calif., and acting president of the Social Science Research Council in New York. He has been dean of the Division of Social Sciences, university examiner and chairman, Department of Education, at the University of Chicago.

The need for a continuing assessment of the progress of education in this country arises from the great demands which are now being made upon education. Most of the goals we seek as a people require education as a means of reaching them. To meet these demands, the American people are furnishing far greater resources for our educational institutions than have ever been available before. Yet much more is being requested. It is clear that the resources required cannot be provided except by using the greatest care in their allocation and use. The public needs to understand more adequately what educational progress is being made and where the critical problems lie on which much greater attention and effort must be focused. The public needs this kind of information in order to give intelligent backing for the decisions that must be made to use resources wisely to produce maximum results.

We have data of this kind about other matters of public concern, such as our population growth, its rate of increase, the extent and direction of migration; about the income levels of our people and the incidence of disease. Need for this information was recognized a generation ago and, over the years, means for obtaining the information were worked out and are continually being refined. We now have useful, comprehensive and comparable data regarding types of morbidity and mortality for various ages, occupations, regions and the like. We know the diseases that are currently the chief causes of death in different age groups, and in different occupations and income categories. We have helpful estimates of production, prices and unemployment ratios. These kinds of data enable the public to understand progress and problems in these fields, and they furnish perspectives from which to make decisions.

But before the advent of the National Assessment, we had no comprehensive and dependable data about the educational attainments of our people. The data available at the state and national level have been reports on numbers of schools, buildings, teachers and pupils, and about the monies expended, but we have not had sound and trustworthy information on educational results. Because dependable data were not available, personal views, distorted reports and journalistic impressions have been the sources of public opinion, and the schools have been attacked by some and defended by others without having necessary evidence to support either claim.

Teaching Johnny to Read

For example, some years ago, a book entitled "Why Johnny Can't Read" had a great influence on public opinion, without any evidence being presented as to how many Johnnies can't read, and in what population groups is there a considerable fraction of non-readers. It turned out that the effect of this arousal of public opinion was to redesign programs for teaching reading largely in schools in which most children were learning to read, rather than focusing the added effort and expenditure on schools where there were serious problems in learning to read. Had data been available at that time on the reading achievements of American children, the public would have had information about the incidence of inadequate reading abilities and could have supported efforts to attack the problems where they were rather than to have stimulated programs that did not reach the schools where they were needed.

Some persons question the statement that the public has not had comprehensive and dependable information about what American children and youth have learned. They know that educational achievement tests have been on the market for 50 years and that they are used widely. Would not the compilation of the scores on standard achievement tests furnish the data the public needs?

The standard achievement tests in common use do not give a dependable measure of what children have learned. They are not constructed to do so. A typical achievement test is explicitly designed to furnish scores that will arrange the pupils on a line from those most proficient in the subject to those least proficient. The final test questions are selected from a much larger initial number on the basis of tryouts and are the ones which most sharply distinguished pupils in the tryouts who made high scores on the total test from those who made low scores. Test questions are eliminated if most pupils can answer them or if few pupils can answer them, since these do not give much discrimination.

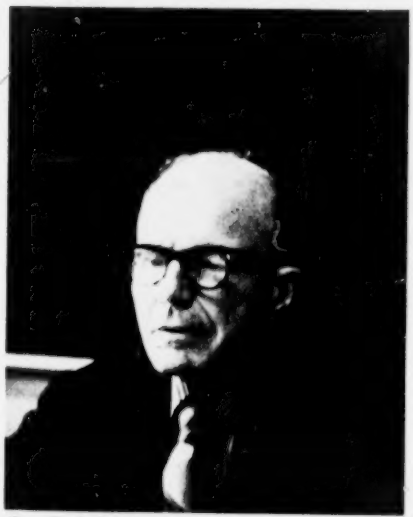
As a result, a large part of the questions retained for the final form of a standard test are those that 40 to 60 percent of the children are able to answer. There are very few questions that represent the things being learned either by the slower learners or the more advanced ones. If a less advanced student is actually making progress in his learning, the typical standard test furnishes so few questions that represent what he has been learning that it will not afford a dependable measure for him. The same holds true for advanced learners.

Schools are discovering that the commonly used standard tests do not show what pupils have learned.

The Score's Not the Answer

This is not a weakness in the test in serving the purpose for which it was designed. The children who made lower scores had generally learned fewer things in this subject than those who made higher scores and could, therefore, be dependably identified as less proficient. Furthermore, a good standard test has been administered to one or more carefully selected samples, usually national, regional or urban samples, of children in the grade for which the test was designed. The scores obtained from these samples provide norms for the test against which a child's score can be related. These tests thus provide dependable information about where the child stands in his total test performance in relation to the norm group. But when one seeks to find out whether a student who made a low score has learned certain things during the year, the test does not include enough questions covering the material on which he was working to furnish a dependable answer to that question.

The National Assessment has been designed to sample the things which children and youth are expected to learn in school, and to find out what proportion of our people are learning these things. The instruments used in the assessment are not tests which give each person a score or a grade. They are exercises that children, youth and young adults are given. Instead of a score, the results are reported in terms of the percent of each population group that was able to perform the exercise. These exercises show the public both what our children are learning and how many are learning each thing. The public is thus able to make judgments about each exercise. How important is this for children to learn? And, in what regions, and other circumstances are most children learning this, and in what circumstances are only part of the children learning?



Dr. Ralph W. Tyler

Schools Must Teach Students, Not Merely "Sort" Them

The purpose of the National Assessment is closely related to the current call for accountability in education. In contrast to earlier years when schools were expected to give major emphasis to sorting pupils so that only a fraction of those who entered school at six years of age would graduate from high school, the discussion of accountability today emphasizes the purpose of the school as learning rather than sorting. It seeks to hold the school accountable for educating all the children, not simply furnishing opportunities for the elite. In meeting the demand that all children learn what the school is expected to teach, data are needed about what is being learned by all parts of the population. Schools are discovering that the commonly used standard tests do not show what pupils have learned nor what proportion of the children has learned each of the things the school is teaching. An increasing number of schools is asking for exercises like those developed in the National Assessment so that they can use them as part of their programs for accountability.

Such instruments could furnish information about the progress and problems of the school in providing the kind and quality of service that the public is now expecting. A constructive dialogue can then be maintained with the community regarding the educational objectives, the efforts the school is making to reach these objectives, the progress pupils are making in their learning, the difficulties being encountered and the steps being taken to overcome these difficulties. Ample information on these matters can make the dialogue a constructive one and can reassure the general public about the integrity of the school in meeting its responsibilities.

In this program for accountability carried on in the local school district, the National Assessment not only furnishes an example of the kinds of instruments to be used to appraise the results of the school's efforts, but the reports of the National Assessment also give a helpful background for the dialogues with the public. The clientele of the local school sometimes places the blame for difficulties in pupil learning upon the teachers and principals without knowing that certain problems are characteristic of the entire nation or region. Because the National Assessment provides background data, the public can gain a broader perspective from which to view the local problems. There will be less tendency to attack the local schools groundlessly because the public will see that most difficult educational problems are not localized and cannot be blamed upon a particular administrator or set of teachers.

In summary, the National Assessment provides a means of helping the public understand the results being achieved by our educational system and the problems encountered. It furnishes a basis for intelligent examination of the situation and helps to identify the places on which to focus more effective efforts.

GEORGE B. BRAIN

Some Values of Assessment

Dean of the College of Education at Washington State University since 1966, Dr. Brain has had a long career as teacher, principal, superintendent of schools, lecturer and professor at some 20 colleges and universities throughout the country. He has served on several federal and private educational agencies, is a past president of the American Association of School Administrators and is well-published in education journals. A graduate of Central Washington State College, he received his Ph.D. from Columbia University.

The assessment of performance is an age-old practice. The last verse of the first chapter of the Book of Genesis reads, "And God saw everything that He had made, and behold it was good." That single act of the Creator apparently established the precedent for a practice of assessing all human behavior. Even though man lacks god-like ominiscience, he tries to follow the precedent, and it was only logical that in time the concept of assessment came to be applied to the field of education.

The process of assessing the educational attainments of students began almost four thousand years ago. The records of the Shun Dynasty of ancient China report an elaborate system of achievement examinations which provided the avenue of entry to the civil service. The practice of assessing the quality of educational attainment flourished even in the Middle Ages. The Jesuits, for example, prepared a detailed and technically sophisticated set of rules for conducting written examinations, which, if followed carefully, would improve many of the examination practices employed in the schools today.

Assessment played an important role in the early development of public education in this country. Horace Mann made provisions for a system of evaluation and reporting in his plan for education for the state of Massachusetts. Other states adopted the practice. When the U.S. Office of Education was established in 1867, one of the duties given to the U.S. Commissioner of Education was to determine the progress of education.

Assessment began to become a science, or at least a developing technology, in the early years of this century. Now with the proliferation of wide-scale evaluation programs, with the perfection of electronic test processing equipment, and with public pressure for accountability in the educational enterprise, assessment continues to play a leading role in the process of education.

Yet, in 1963 when Francis Keppel, then U.S. Commissioner of Education, began talks with a number of educators on the feasibility and need for a nationwide study to describe for the American public the educational attainments of various groups of Americans, the idea of the National Assessment program encountered considerable resistance. National Assessment was confused with a nationwide individual testing program. A number of understandable fears were expressed by those who misunderstood the purpose of National Assessment and who were unsure about the processes and procedures which the assessment program would employ.

The process of assessing performance in education is far more complicated than it appears to many citizens unacquainted with the science of evaluation. As they see it, all one has to do is to give a test. Each student either passes or fails it. The proportion that passes is indication of the effectiveness of the educational program. Where the test comes from, what it really measures, how accurately it tests knowledge, what the scores mean, how the passing score is determined—these and many other tricky questions that trouble scholars in evaluation a great deal

National Asessment provides only part of the information. . . . The balance should come from teachers using tests they have themselves designed.

trouble the general public hardly at all. They trust the technical competence of the specialist in evaluation to do the job that needs doing far more than he ought to be trusted, for even among the scholars in evaluation there remain confusion, uncertainty and sharp differences of opinion. There are, in short, some difficult problems in assessing performance adequately.

Progress and Problems

But, despite these differences and difficulties, the original purpose of the National Assessment project is being realized; namely, to provide the intelligent lay public with census-like data on the educational levels of important sectors of the nation's population in order to furnish a dependable background of information about educational attainment, the progress being made and the problems still to be faced in achieving our national educational aspirations.

As Wendell H. Pierce, executive director of ECS, puts it: "For the first time we are beginning to obtain substantial data on what young people know and do not know; for the first time we have a system by which we can randomly sample knowledge and skills in all sections of the country; for the first time we have bench marks against which we will be able to measure educational change and progress in the future. National Assessment is providing the nation a new and vital resource to be used in evaluating the output of our educational system."

The National Assessment reports which have been published do not contain judgments or assessments about individuals, school systems or states. Instead they indicate areas of strength and weakness in the knowledge, skills and educational attainment of representative samples of the nation's population. In addition, the results are providing the necessary bench marks for the measurement of progress or lack of progress in various subject matter areas over a period of years.

National Assessment has served a valuable purpose in the movement toward educational accountability. It has already made valuable contributions to measurement methodology. It has served as a model for criterion-referenced assessment. Moreover, the first results of the National Assessment program hold promise of providing a tool to improve political decision-making about educational policy at the national level.

Since a degree of uniqueness in schools and educational programs is essential to accommodate the differences in educational needs and purposes among the states, the specific results of National Assessment are not necessarily directly applicable to state educational operations. But the design of National Assessment offers a model for a state assessment plan. The state educational agency is left the task of adapting and completing the assessment system which will be most responsive to the educational needs of a particular state. The National Assessment model serves as an effective guide for the development of such a system. Obviously the state is in the best position to analyze for itself alternatives proposed for its education system; to determine priorities for meeting its own goals, and to develop a quality control program which will ensure economy and efficiency in achieving statewide educational objectives.

Important as it is, National Assessment provides only part of the information needed for a comprehensive program of formal evaluation of student performance at the state and local level. The balance should come from teachers using tests they have themselves designed and built or selected from available standardized instruments. A balanced program of formal assessment should include teacher-made tests, standardized instruments and National Assessment performance-type exercises.

JOHN K. WOLFE

Industry — An Unnoticed Consumer

Dr. Wolfe is manager of University Relations at the General Electric Research and Development Center, Schenectady, N. Y. He has worked with Harvard Medical Center and was chairman of the board of the International Association for Exchange of Students in Technical Experience. Dr. Wolfe was a member of a panel asked to discuss National Assessment Science results at the annual meeting of the American Association for the Advancement of Science in Philadelphia in December. The following are excerpts from his remarks.

The importance of National Assessment to curriculum-makers, textbook publishers, teachers and other educators is obvious.

But there is another great unnoticed consumer who should use the information contained in the National Assessment findings: that consumer is industry.

For example, how might industry react to the recent assessment of science?

Certainly industry will applaud measurement in the science achievement area. We have heard so long that this is a sacred area that cannot be measured and cannot be evaluated. However, until we can evaluate any area, we have no way of knowing its strengths and weaknesses.

Industry will react favorably, for example, in considering assessment results in terms of plant location and employment practices. When an industry considers a certain geographic area for an industrial development, it will be concerned with the educational competence of the people in that area. Important to any industry would be information as to the educational make-up of a community: schooling, general knowledge derived from experience and whether members of that community show a desire for continuous self-education after leaving school. Where but from National Assessment results can this particular information be obtained?

Industries planning to hire locally will want to know the levels of knowledge and skills of their potential employees. One of the important industrial problems now is the development of continuing education programs for people working in plants. To plan adequate and useful continuing education programs, an industry must first find out what employees already know and can do.

If the industry plans to bring employees in from another part of the country, it will be concerned about whether they will be willing to move. A great consideration for any family faced with a move to a new location is the local school systems available for their children. Fortunately for industry, a look at the cumulative data from National Assessment reports will give some indication as to the nature of school systems in different geographic regions plus information on the overall make-up of communities.

It is not by accident that people like IBM are moving into Westchester, New York; that Bell Lab is in Murray Hill, New Jersey; that there are laboratories on the Stanford Campus in California.

Scientific Literacy Low

Industry will also look to National Assessment for information that will help in telling customers about various products. One of the continuing vital problems for an industry is how to clearly describe a product, particularly if it is technically oriented, to potential

The low level of science knowledge of the non-scientific public is appalling. It is not surprising that products are described the way they are.

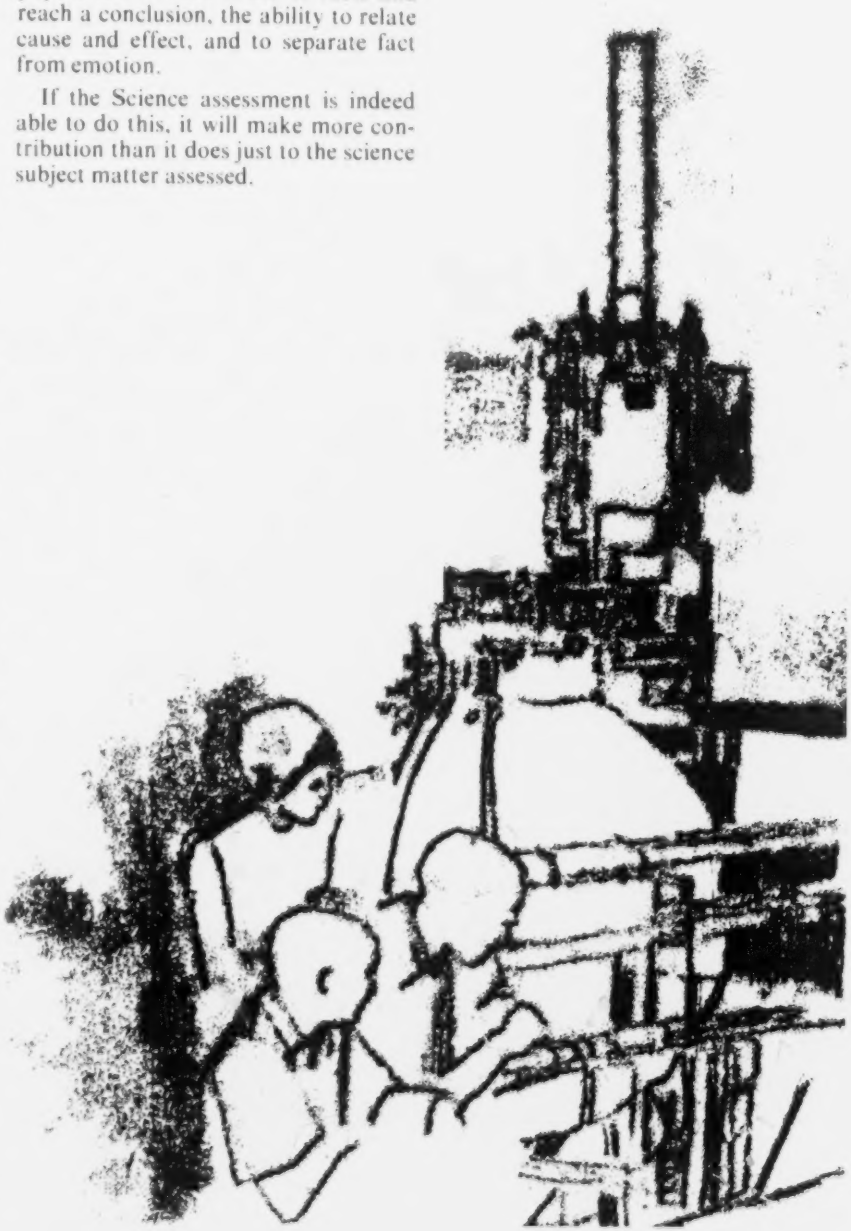
customers. The low level of science knowledge of the non-scientific public is appalling. It is not surprising that products are described and put to the public in the way they are. The level of scientific literacy is probably partly responsible for the kind of advertising we have today.

For example, if I were a manufacturer of phosphate detergent, how would I describe it? I could describe it as having 8 percent phosphorus or 25 percent phosphate. If at the present time phosphate is a bad guy, then I'm going to put on my box that it has 8 percent phosphorus. If phosphate is a good guy, I'm going to try to make it 25 percent phosphate. I'm going to aim it emotionally so that the largest percentage of the public thinks my product is good. If 98 percent of the public doesn't know the difference between phosphate and phosphorus, the manufacturer is going to take advantage of that lack of knowledge.

Some of the responses to National Assessment Science questions can help industry know how to explain vital problems to the public. For example, one of the most difficult areas now for power companies like Con Edison of New York is to interpret to the public the reason for such problems as electrical power shortages. How do you tell the people of a city, who perhaps have been asked to turn off their air conditioners on the hottest day of the year, that they are either going to have to accept the building of an additional nuclear plant or you're going to have to cut the voltage of New York City? Some of the specific data on what people know or can understand about science should help industry in translating information to the public with a minimum of misunderstanding.

I think the name science may be inappropriate for this particular part of the assessment. Of all the categories that National Assessment is using, science may hold the most hope for measuring the logical ability of our population to take a set of facts and reach a conclusion, the ability to relate cause and effect, and to separate fact from emotion.

If the Science assessment is indeed able to do this, it will make more contribution than it does just to the science subject matter assessed.



CARMEN J. FINLEY

Not Just Another Standardized Test

Dr. Finley served as associate staff director and director of exercise development for National Assessment for four years. She left the project at the end of 1971 to become a principal research scientist on the staff at the American Institutes for Research in Palo Alto, California. A former high school mathematics teacher, she has been director of research and data processing, Sonoma County (California) Schools, and has served as visiting professor of statistics and educational measurement at several universities. She received her B.A. from UC Berkeley and M.A. and Ph.D. from Teachers College, Columbia University.

Every year, billions of dollars are spent on education in the United States on buildings, on teachers' salaries, on curriculum planning—but very little is known about the effectiveness of this expenditure. The purpose of the National Assessment of Educational Progress is to gather information which will help answer the question, "How much good is the expenditure doing, in terms of what young Americans know and can do?"

National Assessment, as a nationwide project collecting information about certain groups of young Americans, will, over a period of time, provide valuable information needed to make wise decisions about the allocation of our resources within the field of education.

Concern over the need for this type of national information first began during the time Francis Keppel was U. S. commissioner of education (1962-65). Keppel discovered that in the original charter of the U. S. Office of Education (1867) a charge was given to the U. S. commissioner to determine the progress of education. This provided the initial impetus for National Assessment.

After a number of conferences and discussions initiated by Commissioner Keppel, John W. Gardener, then president of Carnegie Corporation, asked a distinguished group of Americans to form the Committee on Assessing the Progress of Education under the chairmanship of Ralph W. Tyler (then director of the Center for Advanced Study in the Behavioral Sciences, Stanford, California). Their charge was to consider development of an assessment program which would provide benchmarks of educational progress as a basis for evaluating the changing educational needs of our society over the years. Specifically, they were to:

1. Determine how a national assessment of educational progress could be designed;
2. Develop and test instruments and procedures for the assessment; and
3. Develop a plan for conducting the assessment.

Four years of work, financed by the Carnegie Corporation of New York and the Fund for the Advancement of Education of the Ford Foundation, went into defining goals and developing measuring instruments to answer these questions. The work was done in consultation with subject-matter experts, leading educators and interested laymen. Ten subject-matter areas were defined for assessment: Art, Career and Occupational Development, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies and Writing.

In 1969-70 the first actual assessment was made—in the areas of Citizenship, Science and Writing. Last year, Reading and Literature were assessed. Social Studies and Music are now underway and preparations are

National Assessment by design set out to assess what the most capable person could do, what the average person could do, and what the least able person could do at each age level in the assessment.

being made for the assessment of Mathematics and Science next year. By 1974-75, all 10 areas will have been assessed once and several areas will have been assessed twice.

The exercises in each area are designed to measure what groups of people know and can do. Data in each subject area are collected by:

Four age levels [9, 13, 17 and young adult (26-35)]

Seven types of community (inner city,[1] big city,[2] urban fringe, affluent suburb, medium-sized cities, small town and rural, farm rural)

Four geographical regions (Northeast, Southeast, Central, West)

Four educational levels of parents (not more than 8th grade, more than 8th grade but less than high school graduation, high school graduation, some formal education beyond high school)

Color (Black, White, Other)

Sex

For the first time in American education, there is a plan to systematically sample what people know and can do and to report the results to all people involved directly or indirectly in the ongoing process of improving education.

Defining Goals vs. Comparison With an Average

In the National Assessment program specific objectives or goals are defined and exercises are written which determine how well these goals are being met. For example, in Citizenship a major objective is to "Support Rights and Freedoms of All Individuals." One specific way in which a person might meet this goal is to defend the right of a person with very unpopular views to express his opinion and support the right of "extreme" (political or religious) groups to express their views in public.

One exercise which was written to try to tell whether or not this objective was being met is as follows:

> Below are three statements which make some people angry. Mark each statement as to whether you think a person on radio or TV should or should not be allowed to make these statements:
>
> "Russia is better than the United States."
>
> "Some races of people are better than others."
>
> "It is not necessary to believe in God."

This is the goal-oriented approach. The objectives or goals represent a kind of standard which is considered desirable to achieve. The exercises, if they are good measures, tell to what extent the goals are being achieved. This approach tells very specifically what a person knows or can do.

In the norm-referenced approach which is typical of standardized testing programs used by most schools, there are no standards, although the results are sometimes misinterpreted in such a way as to imply there are standards.

The main purpose of the traditional standardized testing program is to place individuals taking the test in rank order from high to low. The results can then be used to counsel or to group youngsters for instructional purposes or to select them for special programs or for college entrance, etc. Standardized tests are also used to evaluate instructional programs or schools or districts or even whole states but the results will only tell whether or not the program (school or district or state) is above or below the average of the group upon which the test was standardized. It will not tell what people know or can do, except in a very limited sense.

Amount of Coverage

The assessment of any subject area for National Assessment is as comprehensive as possible. Every effort is made to measure each of the objectives in each area. The minimal amount of material needed to do this generally ranges from about six to eight hours of actual assessment time. This does not mean that any one person is subjected to so many questions. The total amount of materials is divided into 35-minute units (for in-school administration) and 45-minute units (for administration outside of school). No youngster in school takes more than one unit and adults may elect to take up to four units. Since National Assessment is interested only in group results, it is possible to portion out the materials in this manner and keep the demands placed upon any youngster or school within reasonable limits.

In contrast, the typical standardized test allots approximately 30 to 70 minutes worth of testing time per subject area, and each person takes the same test as every other person. Frequently a number of subject areas are gathered together into a battery, and the total battery may be administered over a period of two or three days.

Type of Administration

National Assessment uses many traditional paper-and-pencil exercises (items), but they are administered in groups of not more than 12 youngsters so that maximum control can be gained over the assessment situation. In addition, all exercises administered in groups use a paced tape, both to in-

[1] Inner city is the 10 percent extreme impoverished area of big cities.

[2] Big city is all other areas excluding the 10 percent extreme.

struct the youngsters and to read the actual questions (except in the area of Reading). Previous research has shown that the usual disadvantage which poor readers or bilingual children have in testing situations can be somewhat alleviated through these methods.

In addition many National Assessment exercises are administered to one child at a time by an interviewer. These are generally exercises of a more complicated nature which attempt to measure the youngsters' thought processes (reasoning and thinking logically, drawing inferences, reaching conclusions, analyzing and synthesizing different points of view) or their ability to perform in some way (i.e. play a musical instrument, demonstrate a scientific principle using apparatus, use a typewriter, etc.).

Traditional testing programs, on the other hand, are typically administered in larger groups, and each youngster must rely on his own reading ability to understand the meaning of the question. Individual testing of students is generally reserved for the small percentage of youngsters who exhibit some type of psychological problems.

Variety of Materials Used and Associated Scoring Problems

Because many of National Assessment's exercises are more complicated than the usual standardized test, the scoring is often more complicated. Standardized tests are generally objectively scored; that is, they can be scored by machine at a very rapid rate. Machine scoring is feasible only if the youngster has a limited number of answers presented to him from which he selects. While National Assessment does use some exercises of this type, many are open-ended, or require the youngster to produce and write out his answer rather than just recognize it from among other wrong choices. In addition, exercises which require him

Basic Differences Between National Assessment and Standardized Testing Programs

There are seven basic distinctions between the National Assessment program and traditional standardized testing programs used by most schools. These differences are summarized in Table 1 and are discussed in the following paragraphs:

TABLE 1

Characteristics of the National Assessment Program As Compared With Traditional Standardized Testing Programs

	National Assessment Program	*Standardized Testing Program*
1.	National Assessment exercises measure how well students as a group achieve desirable goals.	Standardized tests compare students with the average performance of other students.
2.	The time allotted to a given learning area ranges from six to eight hours each.	The time allotted to a given subject ranges from about 30 minutes to 70 minutes each.
3.	National Assessment administers exercises to groups no larger than 12 and to individuals by interview.	Standardized tests are generally administered to total classes or groups of classes in a central location.
4.	Exercises use a wide variety of stimuli and approaches often requiring the student to perform or to provide the correct response rather than just recognize it.	Test items are generally confined to those paper and pencil variety which can be scored by machine.
5.	Exercises are prepared for the "high," the "average" and the "low" ability students.	Items are aimed at the "average" child.
6.	Total scores, which reflect the number of students who got the correct answer, are given to each exercise. People do not receive total scores.	Total scores reflect the number of correct answers a student gives.
7.	Results are reported on an exercise-by-exercise basis.	Results are reported in relation to a standardization group.

to perform in some manner must be judged in a different way. For these more complicated materials very detailed scoring criteria and keys are developed on the basis of field testing which precedes the assessment. Then specially qualified and trained persons are required to do the scoring.

Difficulty of Materials

National Assessment by design set out to assess what the most capable person could do, what the average person could do, and what the least able person could do at each age level in the assessment. Materials, therefore, had to be developed which aim specifically at each of these levels. Results of the first year of assessment indicated that the materials do cover the full range of ability at each age.

In contrast, the typical standardized test best measures within the average range. This is a technical necessity for the purpose it serves—to discriminate among youngsters and place them in rank order. Not infrequently the best students will "go off the top of the test" or the poorest will "fall off the bottom"—thus neither one is adequately measured.

What Scores Are Important

In National Assessment scores are not obtained on individuals. Since any one individual takes only a small fraction of the total amount of materials, scores for individuals would really have no meaning. The scores that are important are those which tell what a group of people did on any given exercise. This gives a kind of item-analysis on a national level and, in keeping with the goals of the program, tells what people know or can do across the nation.

In the traditional testing program items are added and a total score (or set of subtest scores) is obtained for each person. This score is then compared with scores obtained on a standardization group, and it is possible to tell whether the individual ranks high, average, or low with respect to that group.

How Results Are Reported

As indicated above, the results for National Assessment are reported by exercise, i.e., how groups of people perform on each exercise.

The free speech exercise which was given earlier in this article was reported as follows:

Believe a person on radio or TV should be allowed to say:	AGE		
	13	*17*	*Adult*
"Russia is better than the United States."	21%	49%	56%
"Some races of people are better than others."	16%	31%	37%
"It is not necessary to believe in God."	25%	49%	55%
Would allow all three statements	6%	22%	32%

In the traditional testing program reports generally show only the relationship of the student or group to the standardization group. The student or group will be above or below "norm." Since the norm represents an average performance, in effect, the report will show how far above or below average the student or group happens to be on that particular test.

In summary, there are a number of fairly major distinctions which set National Assessment apart from traditional standardized testing programs. This is not to say that one is better than the other, but they do serve distinctly different purposes.

J. STANLEY AHMANN

The First Results

Dr. J. Stanley Ahmann is staff director of the National Assessment of Educational Progress. He was associated with Colorado State University for more than 10 years, where he was professor of psychology, associate director of the Human Factors Research Laboratory, vice president for academic affairs, staff director for institutional research and chairman of the psychology department. He was graduated from Trinity College and received his doctorate from Iowa State University.

Data representing the output of the American educational system was gathered for the first time by the National Assessment of Educational Progress in 1969-70. During this year hundreds of exercises were administered to a national sample of approximately 80,000 young Americans in order to discover what they know and can do in the areas of Science, Citizenship, and Writing. The sample was subdivided into four age groups, namely 9-year-olds, 13-year-olds, 17-year-olds, and young adults (ages 26 through 35).

The three learning areas in the assessment are quite different, one from another. Science is a highly organized learning area that is readily found at many of the grade levels in the typical school program. It is likely that the students in the typical school have been exposed to various aspects of science throughout most of their elementary and junior high school years—and possibly some of their senior high school years. In contrast, Citizenship is much less well defined and is not typically taught as a formal subject-matter area. As some have said, citizenship is "everybody's business and nobody's business." At the same time that it is acknowledged as a highly significant area in the school curriculum, it is difficult to find the intensity of instruction as is so often present in science, with the possibile exception of certain cognitive aspects such as knowledge of structure and function of government. Finally, Writing is one of the basic skills which has an important role in the early years of the typical school program, and presumably permeates many aspects of content areas in the later school years.

What then has been learned about the output of the American educational system as a result of the assessment in the areas of Science, Citizenship, and Writing? To date, eight major reports have been prepared, each containing literally thousands of pieces of information which might well be of value to teachers, curriculum builders, and textbook writers.[1] The principal approach to the analysis of the data has been an exercise-by-exercise approach, consistent with the use of criterion-referenced exercises in the assessment program. National success with respect to each exercise has been tabulated and comparisons have been made between various subgroups within the sample, for example, the two sexes, various sections of the country, and various types of community. In this way a reading on the absolute level of performance as well as the relative level of performance is available.

In the foregoing, it is easy to visualize the great mountain of results which has already been generated by the National Assessment effort. Following are some of the highlights for the three learning areas involved in the first assessment year.

Achievement in Science

Consistent with the principles of good achievement testing, the development of each learning area began with the establishment of the objectives of that area. In the case of science the objectives are classified into three levels in a hierarchial fashion. Four major objectives were used, namely:

Objective I: Knowledge of fundamental facts and principles of science.

Objective II: Abilities and skills needed to engage in the processes of science.

Objective III: Understanding of the investigative nature of science.

Objective IV: Attitudes and appreciations of science and scientists.

Approximately 500 exercises were developed on the basis of these objectives and their subobjectives. They vary greatly in their nature, many of them being of the traditional paper-and-pencil type, while others required the manipulation of apparatus.

The analysis of the Science data has proceeded farther than the analysis of data from any other learning area. In general, it is found that the knowledge of science increases with age, with the exception that the retention of information about science tends to diminish somewhat between ages 17 and young adult when that knowledge concerns aspects of science normally learned in the classroom.

An interesting sex difference appears in the achievement of science. There is a definite tendency of the boys to surpass the girls in this regard. The male advantage is comparatively modest at age 9, increases somewhat at age 13, and becomes even more pronounced at ages 17 and young adult. A study of the Science exercises reveals that the boys

[1] Report 1—1969-1970 Science: National Results and Illustrations of Group Comparisons.
Report 2—1969-70 Citizenship: National Results.
Report 3—1969-70 Writing: National Results.
Report 4—1969-70 Science: Group Results for Sex, Region, and Size of Community.
Report 5—1969-1970 Writing: Group Results for Sex, Region, and Size of Community.
Report 6—1969-1970 Citizenship: Group Results for Sex, Region, and Size of Community.
Report 7—1969-70 Science Group and Balanced Group Results for Color, Parental Education, Size and Type of Community and Balanced Group Results for Region of the Country, Sex.
Report 8—1969-1970 Writing Mechanics: National Results.

do better on physical science exercises than on exercises with biological science content. Furthermore the best performance of the girls was often found in the case of exercises having biological science content. These kinds of findings certainly give rise to speculation with regard to the sex role question, that is, the expectations of society for boys and girls. Are these differences culturally determined in large measure?

Striking differences in achievement in Science were also found in the case of Black respondents. By and large they performed between 12 percent and 16 percent below the national average at the four age levels. On the other hand, Blacks performed best on those science exercises largely dependent upon daily experience and common knowledge. Their poorest performance occurred in the case of those exercises which involved a detached research attitude toward the objects and phenomena of Science.

Illustrative of the great fund of information regarding achievement in Science which is available through National Assessment data, is the analysis of achievement in science by Blacks when each of the four major objectives of Science is considered separately. In Figure 1 is shown the relative performance of Blacks versus the national average for each of these major objectives. Note that performance in the case of Objective I (knowledge of scientific facts) and Objective II (abilities and skills needed to engage in the processes of science) are comparatively low. In contrast, achievement with respect to Objective IV (appreciation of science and scientists) is relatively high. When interpreting data summarized in Figure 1 it should be remembered that the number of exercises for each of the main major objectives varied considerably; for example, the number of exercises devoted to Objective IV is comparatively small in contrast to the number associated with Objective I.

Relative achievement in science was also reported for respondents living in communities of various size and type. For example, it was found that respondents living in highly affluent suburbs performed 5 to 11 percent better than the country as a whole. In contrast, respondents living in highly rural areas performed from 4 to 6 percent below the nation, whereas respondents living in the core areas of large cities showed a 7 to 15 percent deficit.

In view of the fact that the home environment no doubt influences achievement in science, an effort was made to obtain a measure of this variable, at least crudely, by determining the educational level of the parent of the respondent. When the respondents were classified according to the educational level of their parents remarkable variations in achievement were observed. At the four age levels, performance of respondents reporting neither parent educated beyond eighth grade was from 7 to 12 percent *below* the national average. Respondents with at least one parent who attended high school fell from 2 to 8 percent *below* the nation. On the other hand, respondents with at least one parent who graduated from high school achieved between 1 to 3 percent *above* the national average, whereas those of one parent educated beyond high school performed from 5 to 9 percent *above* the national average. Thus we see that a direct relationship exists between the level of performance in Science and the level of education of the parents of the respondent. As the level of education of the parents rose, so did the achievement level in Science of the respondent.

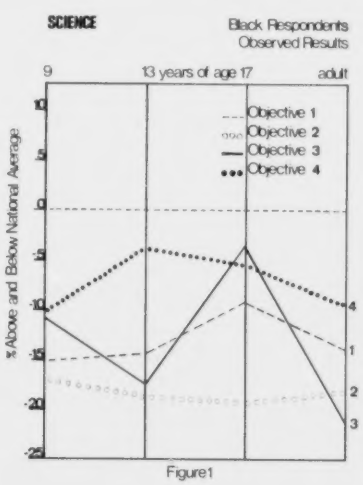


Figure1

Achievement in Citizenship

Achievement in Citizenship is most difficult to define. In an operational sense it can be considered to be the achievement of eight major objectives as determined by various specialists and teachers in this area, as well as concerned laymen. These objectives include such vital goals as: concern for the welfare and dignity of others, support for the rights and freedoms of all individuals, seeking of community improvement through active and democratic participation, helping to maintain law and order, etc. As one might suspect, there is less emphasis on cognitive achievement in this learning area than in Science. At the same time, a greater emphasis on attitudes and appreciation exists in Citizenship assessment than in Science. The Citizenship exercises included both paper-and-pencil materials as well as group discussions.

Again, it was found that there was improvement in achievement between ages 9 and 13, and between ages 13 and 17. A slight decline was found between ages 17 and young adults. This decline did not occur consistently enough to indicate a definite pattern. The instances of lower success for adults can possibly be explained by such factors as loss of skill from lack of practice, or improvements in school programs since the adults received their formal education.

We see a direct relationship between the level of performance in Science and the level of education of the parents.

The difference in achievement in Citizenship between male and female respondents was much less pronounced than that which was found in the case of Science. In general, the differences at each age level were slight with perhaps a slight advantage in the case of young adult male respondents.

Illustrative of the kind of information which can be gained from the analysis of an individual exercise exists in the instance in which stated attitudes of young people towards the acceptance of other races was solicited. The following exercise was presented to the 13-year-old, 17-year-old and young adult respondents:

PEOPLE FEEL DIFFERENTLY TOWARDS PEOPLE OF OTHER RACES. HOW WILLING WOULD YOU BE TO HAVE A PERSON OF A DIFFERENT RACE DOING THESE THINGS?

A. Be your dentist or doctor?

B. Live next door to you?

C. Represent you in some elected office?

D. Sit at a table next to yours in a crowded restaurant?

E. Stay in the same hotel or motel as you?

(For each situation, the respondent had to decide whether he was willing to or preferred not to have this contact with a person of a different race.)

In Table 1 are the data expressed in terms of percent "willing" for 13-year-olds, 17-year-olds, and young adults. Although there is general acceptance of other races in most public situations and relationships, the pattern is somewhat uneven. It should be noted that the acceptance of other races dropped off noticeably when the question is raised of "living next door." For example, about 90 percent of the young adults in big cities were willing to eat together in a crowded restaurant but only about 65 percent were willing to live next door to persons of another race.

Exercises such as the foregoing reveal only the stated attitude of the respondent. It is quite possible of course that the actual attitude of these individuals is somewhat different. Nevertheless, it is gratifying to many that comparatively high percentages of the members of the sample were willing to at least state that they wish to associate with people of other races in a wide variety of public situations.

Achievement in Writing

As in the case of Science, the Writing assessment was also based upon four major objectives. Three of these are writing to communicate adequately in a social situation, in a business situation, and in a scholastic situation. The fourth is appreciation of the value of writing.

Basically, three types of exercises were used: (1) short answer or short essay exercises, where the responses were scored according to whether certain pieces of information were included; (2) multiple-choice questions; and (3) essays requiring writing on a given topic. The essay results were scored for general writing ability, including grammar, word choice, originality, and depth of thought. Later, some of these exercises were also scored in terms of mechanical aspects of writing, such as punctuation, capitalization, agreement, and paragraphing.

Table 1

Stated Acceptance of Other Races in Public Situations and Relationships

(percent "willing")

13	*17*	*Adult*	*Situation*
80 %	70 %	74 %	A. Be your dentist or doctor?
82 %	71 %	67 %	B. Live next door to you?
80 %	77 %	82 %	C. Represent you in some elected office?
82 %	83 %	88 %	D. Sit at a table next to yours in a crowded restaurant?
88 %	85 %	89 %	E. Stay in the same hotel or motel as you?
90 %	89 %	87 %	Willing to associate with a person of a different race in three or more of the above situations
77 %	79 %	77 %	. . . four or more . . .
57 %	57 %	57 %	. . . all five . . .

A sharp difference in performance in Writing existed between male and female respondents.

A sharp difference in performance in Writing existed between male and female respondents. When consideration was given to all of the Writing exercises, the female advantage was pronounced and increased with age. The performance of girls was much better than that of boys on non-essay exercises that require completion of specific writing tasks such as filling out the parts of an envelope or writing an invitation to a class play. Surprising to some is the fact that the results show that the two sexes succeed to about the same degree when writing essays. There is essentially no sex difference in essay performance.

Certain regional patterns were also revealed by the analysis of the data in Writing. For instance, there is a tendency for the southeast to perform below the nation as a whole and lower than any other region compared to the nation. Also, the southeast success decreases as age increases, with typical performance ranging from about 3 percent below the national average at age 9, to about 8 percent below at the adult level. On the other hand, the northeast achievement is highest at ages 9 and 13, but the lead moves to the west at the age of 17, and the central region is highest in Writing at the adult level.

Summary

By way of summary, it is useful to illustrate the manner in which general patterns of achievement in the three learning areas are developing by showing sample figures representing these patterns. Consider Figure 2 which represents the median sex differences in Science, Citizenship, and Writing. The reversal of the male advantage and female advantage in the case of Science and Writing is striking. Given that the dotted line in the center of the figure represents advantage to neither group, then one quickly recognizes that the male advantage increases decidedly with age in the case of Science, whereas the female advantage also increases remarkably with age in the case of Writing. Little difference is observed between the two sexes in the case of Citizenship.

The foregoing is but a tiny sample of the mass of data which was generated by the first assessment year. The full magnitude of the data being produced is difficult to comprehend even when one recognizes that the three learning areas mentioned will be joined by seven others before the first cycle of assessment is complete. Furthermore, steps are already underway for reassessment in the case of the areas mentioned as well as others. In the typical reassessment the objectives will be re-examined and at least 50 percent of the exercises to be used will be new. The remainder will be exercises used in the first assessment but not released to the general public.

Reassessment of Science is scheduled for 1972-73, of Citizenship for 1974-75 and of Writing for 1973-74. Comparisons will then be made with respect to the level of achievement for those exercises which were not released. For the first time it will be possible to determine whether changes have occurred in level of performance in these learning areas. Is the achievement of young Americans in Science, Citizenship and Writing increasing or decreasing? A definitive answer to this question will be of inestimable value to decision makers concerned with education.

Median Sex Differences

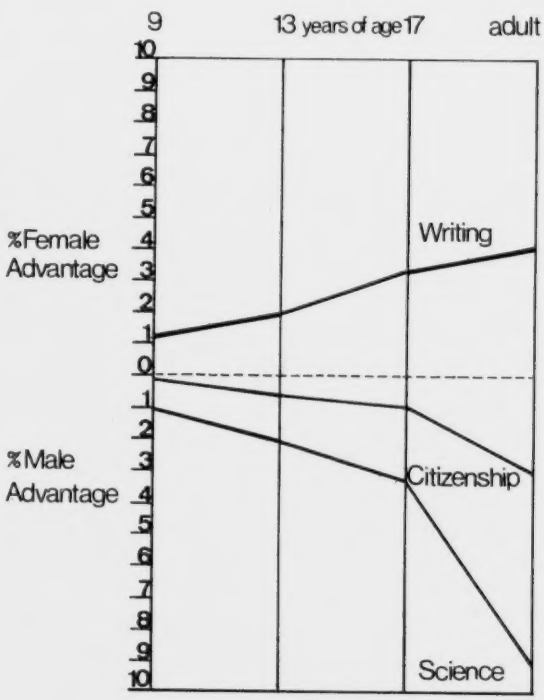


Figure 2

12

Steps of NAEP

Choosing Subjects
to be assessed

Math
Citizenship
Science

Art
Writing
Career and Occupational Development

Music
Literature
Social Studies
Reading

Cycle 1

1969-70 Citizenship, Science, Writing
1970-71 Reading, Literature
1971-72 Social Studies, Music
1972-73 Math, Science
1973-74 Writing, Career and Occupational Development (COD)
1974-75 Art, Citizenship

Cycle 2

1975-76 Reading, Literature
1967-77 Music, Social Studies
1977-78 Math, Science
1978-79 Writing, COD
1979-80 Citizenship, Art
1980-81 Reading, Literature

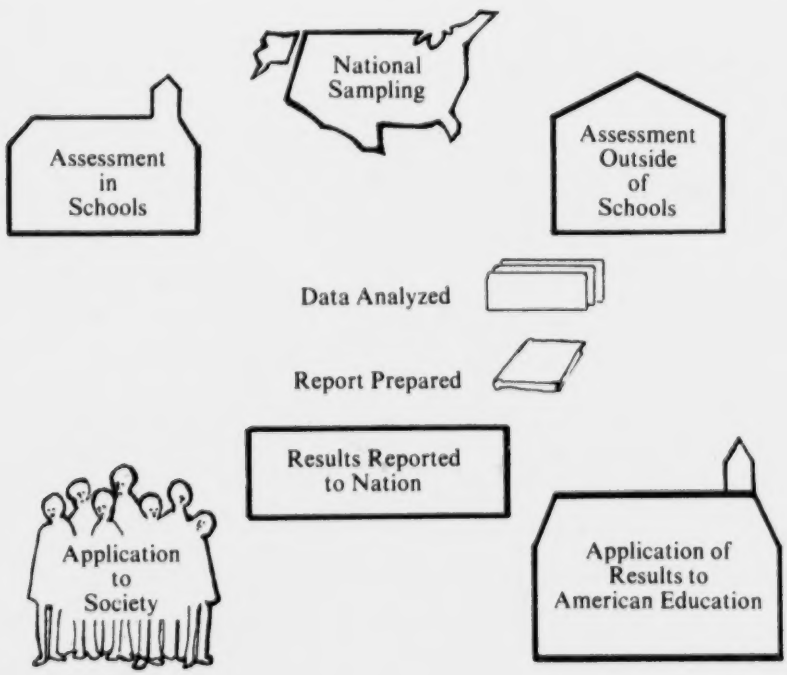
Objectives Developed

Panel

Teachers Scholars Lay People

Exercises Created to Test Objectives

Exercises Tested on Children



19

CLAY ALLISON

Technical Giants of National Assessment

Clay Allison is a nom de plume for a well known free lance writer formerly associated with Time magazine. Mr. Allison is author of numerous books, contributor of articles for national magazines and consultant to many in the writing field.

The Analysis Advisory Committee

Fine Hall, a modern, 14-story tower of brick and stone, stands between Washington Road and Palmer Stadium in the southeast corner of Princeton University's 225-year-old campus. On a Saturday afternoon during the fall, when the varsity is at home, one can go to the top floor of Fine and see as much as a quarter of the gridiron, and if lucky, catch glimpses of the Tigers locked in one of their traditionally deathless struggles.

Most of the faculty and students, whose work-day lives center in the department of mathematics and the department of statistics, which crowd into every corner of Fine, are more interested in such things as multivariable calculus and "analysis on manifolds and differential geometry" or perhaps combinatorics, stochastic process or homological algebra.

On the fourth floor is Room 408, the office of John Wilder Tukey, professor of statistics and associate executive director at the Bell Telephone Laboratories at nearby Murray Hill, New Jersey. The office is plain and simple, offers no amenity other than one overstuffed chair that has seen better days. There is a rack containing miscellaneous books and pamphlets, a table piled high with papers and Professor Tukey's desk, stuffed into one corner and in a state of what might be described as disarray. The room is more of a workshop than an office.

Across the corridor from 408 is a long, rather narrow classroom with blackboards on three sides. On the fourth side the windows look out on a green sward between Fine Hall and Ivy Lane. This classroom often serves as a meeting room for a small group of top-flight statisticians and quantitative psychologists banded together in ANAC, National Assessment's Analysis Advisory Committee, which "puts the frosting on the National Assessment cake."

First organized in 1965 as the Technical Advisory Committee for the budding National Assessment program with Dr. Tukey as chairman, the original committee consisted of Dr. Tukey; Dr. Ralph W. Tyler, a "founding father" of National Assessment and the first director of the Center for Advanced Study in the Behavioral Sciences at Palo Alto; Dr. Lee J. Cronbach, professor of psychology and education at Stanford; Dr. Robert Abelson, professor of psychology and chairman of the department of psychology at Yale, and Dr. Lyle V. Jones, professor of psychology, vice chancellor and dean of the graduate school at the University of North Carolina.

It became the Analysis Advisory Committee in 1969 with Dr. Tukey continuing to serve as chairman and Drs. Abelson and Jones continuing to serve as active members. In September 1970, Dr. Frederick Mosteller, professor of mathematics and statistics at Harvard and vice chairman of the President's Commission on Federal Statistics; Dr. William E. Coffman, professor of education and director of Iowa Testing Programs at the University of Iowa, and Dr. John P. Gilbert, staff statistician at the Harvard Computer Center, became members of ANAC.

Preparing the Package

In its early operations, the advisory

20

committee did such things as give advice on agency contractors who conduct the actual sampling of young Americans in the 9, 13, 17 and young adult (26 to 35) age brackets to determine what they actually know in specific subjects: e.g. science, citizenship, mathematics, writing, etc.

Said Dr. Jones at a recent ANAC meeting:

"We not only gave advice on agency contractors, but we even designed exercises. There was no one else and no other agency to do the work. We continued in this way until the exploratory work was over, and it was assured there would actually be a national assessment program.

"Then the sampling schemes had to be worked out, the exercises had to be selected and we had to decide on the form and organization of the reports. We had to settle on the definitions of size and type of community, educational level of parents, etc. Finally we had a package to sell. Ralph Tyler and others then spelled it out, funding followed and National Assessment became operational, complete with procedures and staff."

Dr. Tyler, a bouncy and vigorous near-septuagenarian, thinks National Assessment is forging ahead "despite the fact that we were delayed at least a year by the job of getting proper exercises. The exercises at the outset did not really sample nor did they give us enough information. The need to develop exercises that more faithfully reflect the objectives of the school is still with us. The second critical problem is how to report these data in ways that will be most meaningful and proper."

Although Tyler sits in on ANAC meetings only occasionally these days, the committee's objectives are in line with his thinking. Chairman Tukey and his associates see their principal jobs as analyzing data and deciding on how it should be reported.

The Heart and Guts of NAEP

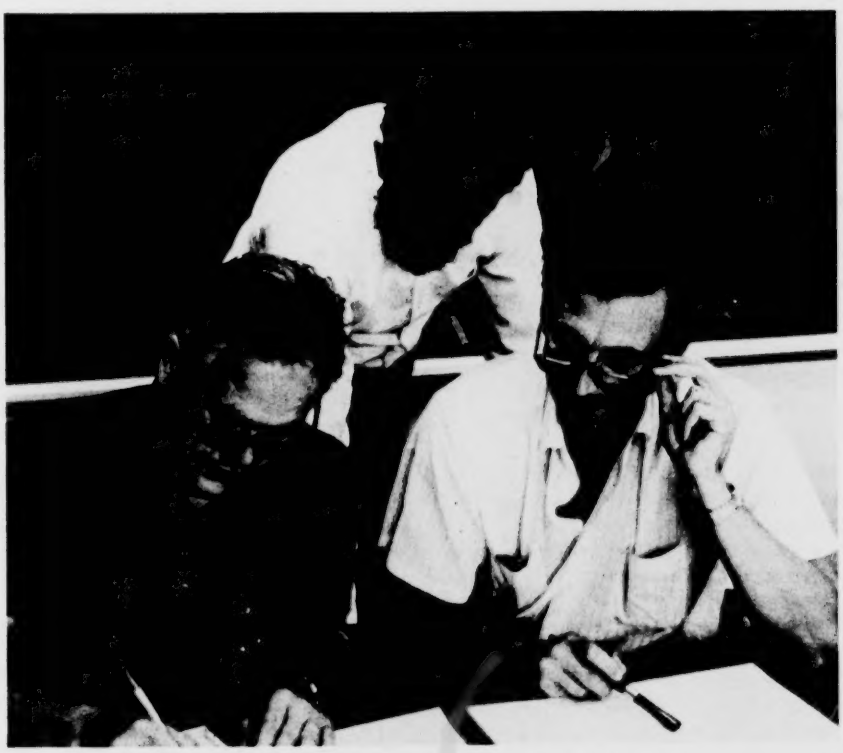
The committee wrote Science Reports 1, 4 and 7 which reported the raw data of science sampling in 1969-70 on the basis of sex, region, race, parental education and size and type of community. While ANAC did not prepare the reports issued thus far on citizenship and writing, it reviewed them before they were published by NAEP.

"The general strategy today is that the committee's contribution in its present form will be complete when we finish the science reports," said Dr. Lyle Jones. "We feel that as consultants we are doing too much operational type work in analysis and reporting. We have had the major responsibility in science, but when we have finished with it, it should serve as a model for the other categories. In the future ANAC will be much more advisory and will leave the analyzing and writing of reports to the National Assessment staff.

"Essentially what we have done is take the responsibility to clear up the methodological situation and try to find out how to analyze things so they make good sense. Our experience has been that if they make good sense in science then they make good sense in other areas as well."

Said Dr. Abelson:

"There are many problems in an assessment program, but among the



Hard at work . . . long into the night

basic ones are these: preparation of adequate and proper exercises to be administered to selected groups (in the case of National Assessment: to 9-, 13- and 17-year-olds and to young adults, age 26 to 35) on a nationwide basis. The first question is, will the exercises do the job? Once the sampling has been done and the exercises scored a check must be made for errors of omission and commission. Then they must be analyzed and reports written. This is where ANAC comes in and is why it might be referred to as 'the heart and guts of National Assessment.' "

Figures, results and scores, standing by themselves, don't mean very much or may be actually misleading. Professor Tukey has a favorite story to illustrate the point:

"In this community there were two hospitals doing the same kind of anesthesia and surgery. The statistical results at one hospital were much better than at the other. Judging from the figures alone one had a good record in administering anesthetics and performing surgery, the other fell short. But study showed that the one with the good record did most of the 'easy' jobs while the other handled the tough emergencies brought in off the street and the aged poor. When these factors were taken into consideration in analyzing and balancing the statistics, the records of the two hospitals were much closer."

Chairman Tukey of ANAC

Overcoming Biases

Judgment, then, must be used in handling and interpreting statistics along with the mathematical tools always used by such master statisticians as the members of ANAC. The process of "balancing the data" was created by ANAC to assist in judging the results properly.

"Sometimes you have to make special fixes for various kinds of troubles that show up in statistics, and when done correctly, they may very well be the best thing to do. In handling data we must always determine if a number used one way is going to be more informative than if used in another way. So you make a judgment. You always have to wind up using judgment in analyzing data, and we hope that whatever biases the members of ANAC may have are sufficiently diverse so that we get good answers and good judgment. Nobody is without biases . . . and under the right circumstances the world seems to do pretty well in canceling biases out. That is clearly what we want to do in analyzing National Assessment data. It is secondary, it seems to me, to the exercise of good statistical inside knowledge and all the sort of thing that goes into doing a good analysis, but one has to include judgment."

All of the members of ANAC have professional careers in full swing at their respective universities and some of them hold down several posts, but they are very busy consultants for NAEP. They meet an average of eight to ten times per year and each time spend a couple of days and nights pouring over the assessment data, arguing about what it says and what it means, and writing reports.

Since they all have demanding schedules it is no mean feat to get them together from five scattered states. About half of the meetings are held in Tukey's

"Let's marinate on that for a time."

Princeton aerie, the remainder in Palo Alto or Washington or Chicago or someplace else.

John's Team

When they do get together they organize quickly, lay out the ground rules for the meeting and get right to work. There is no formality, no pulling punches, no hanging back. Each man is an expert in his field (statistics, psychology, psychometrics, educational measurement), each is aware of his command of his subject and of the outstanding professional capabilities of his fellow committee members. Most of them have been associated at one time or another with the Center for Advanced Study in the Behavioral Sciences at Stanford University in Palo Alto. They have worked together for a long time, pull together easily as a team.

The committee uses a jargon of its own coupled with terminology from their respective disciplines and the bright new world of data processing. In addition to the all-committee terminology, Chairman Tukey has some favorites of his own, some of which can be as difficult for a visitor or a new member of the committee as the committee's acronyms. When Dr. Tukey wants his committee to think about some debated point he is apt to say "Well, let's marinate on that for a time." And when he says, "Illuminate me" he wants a more detailed explanation. Another Tukey favorite: "I hear your words, but I'm not sure what you mean."

A strong, stockily built man with sparse, collar-length graying hair, Dr. Tukey's conduct of a committee meeting can be quite a show for the uninitiated. He favors black, short-sleeved knit pullovers, suntans and Keds, uses his horn-rimmed glasses to do paper work, stare over and as a sword for conversational thrusts, slashes and parries.

He has a round, sun-tanned expressive face. He peers, stares, grins, and when making an important point, often screws up his right eye until it is tightly shut.

Cardboard boxes and a long table

Another Tukey habit which can startle, amuse or annoy strangers and visitors is his custom of doing paper work while conducting a committee meeting or carrying on a person-to-person conversation. He typically runs an ANAC meeting for two nights, a day and part of a second day, says "I am of the opinion, and I think it's right, that the proper duration of a committee meeting in terms of its effect is not the number of days, but the number of nights. When we have two night meetings instead of one I think we get more done. Our meetings have a nasty habit of being on weekends . . . (long interlude) . . . which doesn't fit well into anybody's schedule . . . (long interlude) . . . though we have a certain amount of doing together."

Peanut Butter and Prunes

The meetings usually start Friday night, continue Saturday and Saturday night and then go for awhile on Sunday.

A Tukey-chaired meeting is apt to be a long meeting. He thinks in terms of a half hour break for dinner, may arrive for a meeting that starts at 5 p.m. equipped with a jar of peanut butter and a box of crackers, keeps the committee going on and on into the night. The members gradually sink lower and lower in their chairs, may wind up sprawled on two or three chairs. Tukey sometimes stretches out on his back on two or three chairs, locks his hands behind his head and conducts the meeting while staring at the ceiling or with closed eyes.

He is also a prune man. At some point or other during ANAC gatherings cellophane bags of ready-to-eat prunes appear. Tukey eats them thoughtfully and puts the pits on the table in front of him, in a mathematically straight line. So do his colleagues.

While this sort of thing might lead an outsider to believe that ANAC meetings are fun and games, they are no such thing. Chairman Tukey and his committee members are highly professional, dedicated men doing a serious and important job for NAEP because

Dedicated men doing a serious and important job for NAEP because they believe the program is important to education and the nation.

they believe the program is important—to education and to the nation.

Fifty Pounds of Data

At a given meeting each member, including Chairman Tukey, will have a specific assignment. For example—at a meeting in Princeton to work on the results of sampling young people on their knowledge of science, each member was asked his preference, then assigned to interpret the data in a specific category and write a report: Dr. Coffman to regions, Dr. Abelson to race, Dr. Jones to education of the young peoples' parents, Dr. Gilbert to communities. Dr. Tukey took the one remaining: sex.

They worked at a long table with a cardboard box containing fifty pounds of data set out for each man. The material, after it had been gathered from the field, had been scored and put on tape by the Measurement Research Center in Iowa City, Iowa, then forwarded to the Princeton Computer Center. The computer run, which required the services of 10 technicians, had cost $14,000.

The discussion went on for the entire weekend, will be a continuing one for ANAC meetings in the months ahead.

The general procedure is that once they have a meeting of the minds on a specific batch of data, each man writes his assigned report, then circulates to the others. The next step is a conference call in which the entire committee participates. During the call, which is usually at night and continues without interruption for five or six hours, the members edit each report, word by word, line by line, paragraph by paragraph.

After each chapter is approved by the committee, Tukey writes a summary which gets the same treatment. The chapters and summary then become the technical report which is published by National Assessment of

Charting the data

Educational Progress and is made available to educators and others who have an interest in the assessment results.

Dr. Coffman, who, with Dr. Gilbert, is one of the newer members of the committee, has a slightly different view of ANAC meetings and procedures as a result of his long identification with traditional type testing programs at the University of Iowa. The traditional test compares students with fellow students, whereas National Assessment makes no such comparisons, instead seeks to learn what students actually know about specific subjects.

Said Dr. Coffman:

"I've been interested in looking at data from a new point of view, though I must say I'm amused by some of the criticisms made of traditional testing. Today I hear people blaming reading specialists because children can't read, but that simply isn't true.

"But the effort to look at testing data from a new perspective and summarizing the material item by item will result in new ways.

"Actually the individual teacher (already) has more information than she can cope with just as our analys's committee has more information than we can really handle.

"However," Dr. Coffman continued, "the important thing about the way we are doing National Assessment is that no individual gets more of a test (sample or exercise) than he can handle, so we can administer more complex items in an hour's time. We can set up laboratory experiments. All this gives us more flexibility in what we measure; it gives us the kind of information we need to make general judgments.

"A teacher does this in his or her mind for his or her pupils, but you can't very well summarize what is in the heads of thousands of teachers.

"The NAEP ability to generalize because of careful sampling and widespread cooperation is a great asset. National Assessment didn't ask for much time and as a result got better cooperation from the schools and a better sample than anyone ever got before. Not the biggest, but the best," he concluded.

Some members of ANAC are beginning to wonder, after their tough years as an operating body, if they will be able to find the same challenge in a purely advisory role, but all are anxious to give it a try.

The members of ANAC are not only former members of the Center for Advanced Study in the Behavioral Sciences at Palo Alto "club" but most of them have worked together on other projects. For example, Tukey, Ableson, Jones and Gilbert have worked together for NBC Election Returns.

But, happily for NAEP and ANAC, there is still another bond: Said Mrs. Tukey:

"They are all John's friends."

They are important friends of the National Assessment project, also.

ZOE VON ENDE

Layman's View

A graduate of the University of Wisconsin, Zoe von Ende has been a staff writer with the Denver Post for the past ten years. A former women's editor, she is currently doing feature writing for the Post.

My experience with National Assessment has been limited to only two encounters: participating in a writing review conference, headed by former U.S. Senator Wayne Morse, and through a feature story I wrote dealing with sexual differences in learning as revealed in NAEP data on science, writing and citizenship. Presumably, my role as parent helped qualify me for both assignments.

Consequently, my feelings and impressions about National Assessment are ambiguous. I find much that I like; but there are some aspects I question, such as:

1. I am most favorably impressed with the people associated with National Assessment. Everyone I have met, mostly staff members in Denver, are thoroughly professional in their approach to their jobs, highly motivated, flexible and objective.
2. I am equally impressed with the thoroughness that has gone into the selection of exercises. Other technical aspects of the program, sample-taking and administering, for instance, appear to be well done, too.
3. The broad base of participation from citizens evidently is good, also. It may slow the program, but I think it's basic to the success, particularly when it comes to eventual presentation of material to those most political people, local school board members.
4. Great effort is being made to take into account socioeconomic differences in individuals being tested. The criticisms registered by my panel often, maybe usually, dealt with our finding middle-class bias in the exercises. We were most emphatic every time we came upon this, stating our opinions and reasons for them.

By Whose Standard?

At this point, though, I come to the point of my negative reactions to National Assessment. They include:

1. Removing that middle-class bias is so difficult, maybe impossible. My greatest concern here is that cultural values inferred from the exercises seem to me to be on an absolute, not relative, basis. By your or my standards, a 13-year-old, for instance, writes poorly, yet she definitely conveys a message. The message might be that she is totally unable to articulate her feelings or observations, yet it is a message. Children in the group we looked at did not react in the expected manner to Cannonball Adderley's version of "Mercy, Mercy." However, they told me one important thing: they didn't relate to *that* particular music.
2. Like so many others, I am concerned with eventual utilization of National Assessment results. My biggest concern is that teaching methods will be largely unchanged; that change, if any, will be on emphasis of subject matter. It seems to me that many learning difficulties lie mostly in the realm of attitude, motivation and methods of schools rather than in subject matter per se. Questions I ask myself about National Assessment are: Will it, or can it, stimulate teachers and schools generally to innovate, to stimulate learning, to develop the stamina so necessary to establish individualized goals, to accommodate the system to the student rather than vice versa? Can National Assessment help schools get rid of the right-answer syndrome and replace it with learning how to learn?
3. The National Assessment questions that we reviewed, like practically every other test I've ever had anything to do with, did seem to test ability to answer questions more than ability in a given area. Is this a built-in weakness?

Publicizing National Assessment seems to me to be an important and difficult job. News from National Assessment isn't startling and headline-grabbing. It's more likely to be quiet and significant, not loud and significant. And, as I said before, acquainting local people, political leaders especially, with the whole project is essential to its basic purposes. It therefore must compete with hundreds of stories for space in the local paper or time on local television programs, assuming National Assessment wants its message spread to those overworked, nebulous grass roots.

Comments and Quotes

PRESIDENT RICHARD NIXON
special education
message to Congress
March 3, 1970

"The greatest need in the school systems of the nation is to begin the responsible open measurement of how well the educational process is working. It matters very little how much a school building costs; it matters a great deal how much a child in that building learns. An important beginning in measuring the end result of education has already been made through the National Assessment of Educational Progress being conducted by the Education Commission of the States."

ELIZABETH WOOD
scientist, formerly with
Bell Telephone Laboratories

"I am concerned about what we are not doing for the students who somehow have failed to reach the objectives set by National Assessment. . . . To me, the National Assessment results must be very carefully studied for their implications for the way in which science textbooks are written and illustrated."

SIDNEY P. MARLAND, JR.
U.S. commissioner
of education

"For the first time in our history, we are getting a valid picture of what people have learned from formal schooling. The reports provide a scientific base for measuring the nation's educational progress or decline, and gives the public a look at how its educational investment is paying off."

RALPH W. TYLER
chairman of
Exploratory Committee on
Assessing the Progress
of Education

"The National Assessment is designed to furnish information to all those interested in American Education regarding the educational achievements of our children, youth and young adults, indicating both the progress we are making and the problems we face. This kind of information is necessary if intelligent decisions are to be made regarding the allocation of resources for education purposes."

(National Assessment of Educational Progress, Science Objectives, 1969)

TOM McCALL
Oregon governor,
former chairman of the
Education Commission of the States

"National Assessment will give us gauges that will help school systems in their programs to restructure curricula. For the first time the people of the U.S. will be able to determine how well their educational dollars have been spent."

ROBERT E. STAKE
professor of education,
tests and measurement,
University of Illinois

"National Assessment has been designed to give us a graphic plot of progress through time in different geographical areas. Developing the time dimensions will try our patience because, according to the present National Assessment schedule, it will be at least six years before three points can be plotted as the beginnings of a trend line. . . . National Assessment today is at the beginning of a massive, expensive field trial; a reasonable evaluation of its utility cannot be made before 1975. We can take some comfort . . . in the fact that its staff is honest, competent and productive."

RUSSELL VLAANDEREN
ECS director of research

"National Assessment of Educational Progress is a pioneering project, both from the standpoint of goals to be achieved and the techniques used. New administrative and analytical techniques have been developed which will, no doubt, cause controversy within the profession. However, we will know more about what our children know than at any other time in the history of formal education."

ROBERT L. EBEL
professor of education and psychology
Michigan State University

"Interest in education has never been higher in this country. Neither has the cost of providing it. Both of these generate concern for quality. . . . Education is no longer primarily a matter of local interest. Support and direction from state and national governments have become increasingly influential."

NEW YORK TIMES

"The recent revelations, contained in the first report by the Assessment, could have been so watered down by generalizations that they offer few hard clues to the fluctuations in reading ability, mathematical competence and other basic indicators of educational success. It would be unfair to write off the validity of the Assessment itself on the basis of this first, disappointing report. The question is whether the Assessment is prepared from now on to be more specific in its efforts at comparison, no matter how odious they will seem to those whose professional performance or general support of education may be found wanting."

THE WASHINGTON SUNDAY STAR

"What the Kinsey Report was to American sex, the National Assessment may be to American education."

FRANK H. WESTHEIMER
chemistry professor, Harvard University
member of the President's Science Advisory Commission

"The National Assessment should be looked at as policy research. That is to say, those items should be stressed which can have influence on public policy. For example, one of the variables which was not mentioned or worked on in the assessment is the variable of curriculum. No one asks whether the items that people do or do not know are items which are or are not taught. But if one is looking at what should be done in the schools, one should ask the question whether the things which are taught in the schools are, in fact, learned."

SCHOLASTIC TEACHER

"ECS and the NAEP research staff offered no interpretations of these first results and provided others few leads on which to base their own conclusions. Their position is that they do not want to seem to be passing judgment on American education or educators. Consequently, both teachers and laymen may have some difficulty in deciding which findings they should be pleased about and which ones they should view with alarm."

FRANK B. WOMER
former National Assessment staff director

"The ultimate goal of National Assessment is to provide information that can be used to improve education at any and all levels where knowledge will be useful about what students know, what skills they have developed or what their attitudes are."

ADDISON LEE
science education professor
University of Texas

"I think it's important to keep in mind that this study (the NAEP Science Report) provides information rather than answers to many questions that we might pose. If we think of it as information, then we think of it as a pool from which we can study problems. And if we study (the problems) with the proper kinds of constraints and recognition of the limitations which they have, then we may be able to use this information in a much more effective way. I think it's quite clear that analysis of the responses, and the nature of the responses and the various groups and comparisons that can be made, will be extremely useful to curriculum makers."

U.S. NEWS & WORLD REPORT

"As the National Assessment of Educational Progress continues, Americans will learn a lot more about the knowledge and attitudes of their young people. And the first indications are that the findings may give parents more concern than comfort."

JAMES A. HAZLETT
administrative director
Education Commission of the States

"National Assessment is a bold new education product measurement program and is consistent with the mood in the country to determine the yield of the educational dollar."

DENVER POST

"These assessments will give not only educators, but any concerned laymen, a new and valuable basic set of facts by which they can judge how effective a job our educational system is doing."

FRANK B. WOMER and MARJORIE M. MASTIE

Can National Assessment Change American Education?

Dr. Frank Womer was staff director of NAEP for four years (1967-71). He recently resumed his professorship at the University of Michigan, where he also has served as consultant in testing and guidance in the Bureau of School Services. He holds a Ph.D. in educational measurement from the same university. Marjorie Mastie was an assistant to the director of exercise development for NAEP and currently is serving as a consultant for the organization. The accompanying excerpts are from an article printed in the October 1971 issue of Phi Delta Kappan.

A recurring concern, both among those who support National Assessment and those who have reservations about it, is the ultimate utility of the results. How will they affect education in this country? This is a very difficult question. While National Assessment is designed to provide general information, it is not designed to produce answers to specific educational questions.

Certainly the originators of National Assessment expected the project to contribute to improved educational decision making. Certainly they felt that better answers can be produced by decision makers if they have more information. Their thesis was that someone needed to begin, systematically, to gather information (not answers) about educational outcomes in this country.

These originators were men and women of sufficient vision to see innumerable possibilities for the use of assessment-type information: for legislators faced with hard decisions; e.g., whether to allocate extra monies for instruction in reading if it means refusing requests for other educational needs; for school board members faced with the question of how to deal with educational needs of disadvantaged groups or minority groups; for curriculum specialists and teachers faced with decisions of how best to allocate class time to educational materials related to specific goals.

But National Assessment was designed to be just one information-gathering project, to fill one information void. Well-designed state assessments, local school district assessments, and special research studies seeking answers to specific educational questions will all be necessary to complete the picture.

Utility of Assessment Results

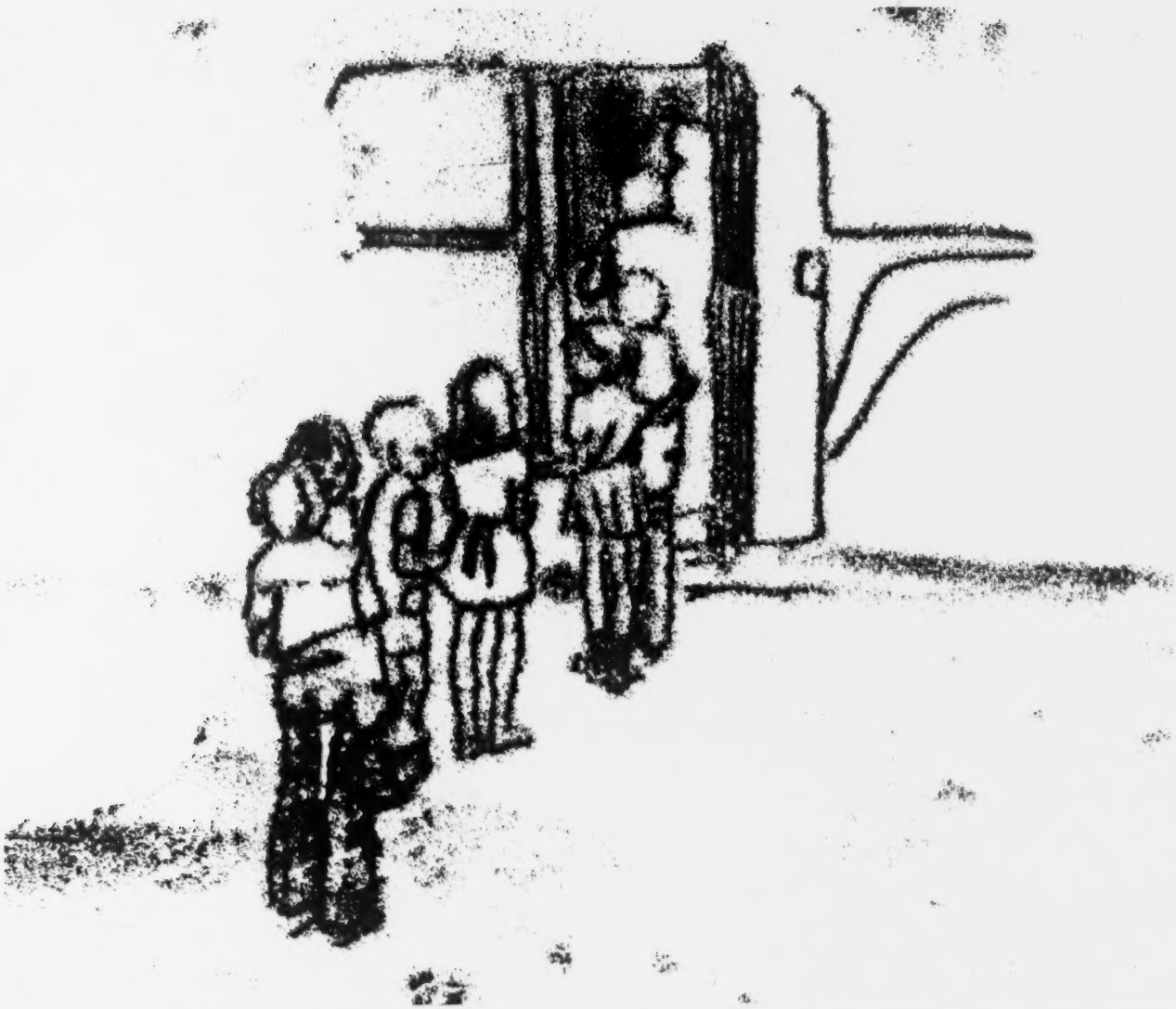
The original and basic purpose of the National Assessment of Educational Progress (NAEP) project was to assess change in knowledge and skills that relate to important objectives of American education. This purpose cannot be met for some years, not until a subject area is assessed the second time.

One of the policy committee members several years ago expressed the fruitless hope that, ideally, NAEP would collect data for its first complete cycle before releasing any results, then reveal "before and after" comparisons.

This was never possible. Yet both supporters and critics of National Assessment are concerned that something should "happen" immediately to demonstrate the utility of the project's results. "What can National Assessment do for me now?" they ask.

Because assessment is a project designed to be of utility to many different people but is not designed to answer specific questions, it is difficult to speculate about actual outcomes. Findings have not been available long enough to point to specific situations in which a legislator or a board member or a superintendent or a teacher has used the results.

National Assessment can serve and has served as one of many stimuli in the movement toward educational accountability. It can make and has made contributions to measurement methodology. It is beginning to release results that have considerable imme-



diate utility for curriculum evaluation. As soon as all science, writing and citizenship reporting is complete (mid-1972*), curriculum specialists will have a wealth of data that should be examined, exercise by exercise, for cues as to which objectives in a given area need more (or less) attention.

For example, National Assessment might show that factual knowledge in science seems to be covered much better than problem solving. A science educator then might want to consider how to beef up instruction in problem solving. He might even want to try some of the exercises on his own students, informally.

Or assessment results might show that while the Southeast Region results are consistently below the national average for citizenship, Southeastern adults score higher than the nation on exercises related to the objective, "Help and respect their own families." A

*Science, writing and citizenship "national" results are available now from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

24

If I were a member of a board of education, I would want to inquire whether our students were studying the issue of freedom of expression.

sociologist might want to investigate the reasons for such a reversal.

Again, National Assessment might show that, while black-white differences are sizable for some science exercises, they are small for others. If such a result appeared it would behoove educators to compare the features of "big difference" items with "small difference" items for explanations of specific knowledge differences that might be masked in a single overall score.

Legislators and board members probably will be attracted more strongly to the citizenship and reading results. One citizenship exercise demonstrated clearly that most young people and young adults do not believe that the government should allow certain very controversial statements to be aired on radio or TV. If I were a member of a board of education and saw that result, I would want to inquire of my superintendent whether our students were studying the issue of freedom of expression, freedom of the press, etc. After reading the total citizenship report, I might even be disturbed enough to suggest that additional resources be allocated to citizenship education, or that our director of curriculum be asked to appear before the board to discuss our practices in citizenship education.

If I were an educational researcher, I might well find myself intrigued with some of the analyses of writing results that have *not* been done. Several writing exercises required respondents to write essays about given topics. These essays are scored by teams of readers (English teachers). A number of researchers have tried another way of scoring essays, using computers. Perhaps I should seek access to assessment data, secure some foundation funds and determine whether computer scoring of essays could just as well provide the information needed.

If I were a linguist I might seek access to a sample of essays from each of the four regions, in order to analyze detailed differences that may not appear in an overall "score" for each essay. Or I might secure the voice tapes of students responding to a literature exercise in order to study speech patterns from this national sample. (Currently, National Assessment data are available to others on a limited basis. Planners are now considering how and when to develop an NAEP data bank.)

If I were a specialist in school finance I might wonder why the NAEP has not chosen to break down its results into two or three levels of financial support for schools and report them accordingly. If I felt this strongly enough, I would communicate with the policy committee, through the Education Commission of the States. National Assessment must be responsive to the needs of educational decision makers. It is a tool which will be effective only if widely used.

Cause and Effect

National Assessment is designed to describe levels of achievement but not to ascribe reasons. If the results indicate that knowledge in the physical sciences seems to be less widespread than knowledge in the biological sciences, National Assessment will not tell you why. Perhaps the fact relates to what is taught in the schools (more students take biology than physics); perhaps it relates to greater dissemination of biological information through news media, TV, etc.; perhaps it relates to greater motivation of students in one direction than another.

If the results of National Assessment indicate a decided lack of knowledge of the structure of government, particularly at the state and local level, who can say whether it is the fault of our schools, of other social agencies or of a general lack of interest on the part of many citizens in their government?

School Comparisons

National Assessment is designed to provide information about the performance of large, well-defined groups (regions, sex, etc.) but not to make state-by-state comparisons, let alone school-by-school comparisons.

National Assessment has a clearly defined role at the national level, but it is not *the* information-gathering project for all education. Certainly every state and every school district should have its own assessment-evaluation procedures. Some of these procedures may be like National Assessment, some may not. Each set of procedures should be directly related to the information-gathering goals of the particular state or school district.

The educational and political needs of the states may well lead to some involvement of the NAEP at the state level. The role of the governing body for NAEP—the Education Commission of the States—is to advance education at the state level. There certainly is a potential utility, and the early fear of state comparisons seems to be much less now. A compromise between no state involvement and every-state involvement would be a voluntary approach. If the National Assessment model happens to fit the goals of a given state, it could tie in with the NAEP. If the goals overlap somewhat, a state might profitably use certain aspects of the national model. But if the overlap seems minimal, particularly if a state wants district-by-district comparisons, it would be much better advised to develop an assessment geared to its own information needs.

Conclusion

The speculations made here are intended to illustrate the innumerable areas in which National Assessment results may prove to have considerable utility. The search for what assessment will tell us and *what we should do about it* is only beginning. The ultimate success of National Assessment depends upon teachers, administrators, board members and legislators using results to improve their own decision making.

GEORGE H. JOHNSON

Making the Data Work

Currently associate staff director of National Assessment, Dr. Johnson was affiliated with Rand Corporation and the System Development Corporation for 12 years and has been director of the Institute for Communications Research at the American Institutes for Research in Washington, D.C. He received his doctorate in educational psychology at Syracuse University.

"Is anyone using it?" is an increasingly frequent question from National Assessment audiences. Since the eventual success of National Assessment must be measured in terms of its impact on educational practice, content and decision making, this, of course, is very appropriate.

Before answering this question, however, it is necessary to identify what there is about National Assessment that is capable of being "used." National Assessment "products" are of two basic types. The first of these consists of the objectives, procedures, criterion-referenced exercises, sampling plans, etc., which together provide a model, technology and materials for conducting assessments at state and local levels. The second applicable output is made up of data on achievement in each assessed subject area, which can be analyzed and reported in various ways. The potential for "use" or application, and the audiences (or users) involved, may be quite different for these two kinds of National Assessment "products."

On December 2, 1971, the ECS Steering Committee approved the recommendation of the NAEP Policy Committee that an Office of Utilization be created in response to the need to effectively use and apply the results of NAEP. Planning is currently under way for establishing several projects aimed at furthering the utilization function.

Furthest along in development is the use of the model, technology and materials developed by National Assessment. Available are: (1) statements of objectives for each of the ten subjects under assessment, with procedures for

Special projects such as workshops for state and local assessments are in the planning stage.

achieving agreement among specialists, educators and representatives of the lay public; (2) methods and guidelines for developing criterion-referenced exercises to measure the objectives; (3) a sampling design and administrative procedures for implementing the design; (4) a selection of the actual criterion-referenced exercises used in earlier assessments. In this area, NAEP is trying to identify present applications, such as, use of the objectives in local goal setting, adoption of the NAEP model for state or local assessments and administration of the released exercises at a state or local level to generate comparative data. Several such applications are known to be in progress, and NAEP is rendering assistance in several of these areas. However, the extent and range of such current applications need to be determined to aid in planning future utilizations.

In addition, materials are being prepared to inform potential users about NAEP and to provide the knowledge and materials needed for assessment applications. Materials include modifications to the National Assessment model which is adaptable to state and local circumstances, development of cost models and unit costs for the National Assessment approach and making released exercises in subject areas available to potential users.

Still further, active dissemination efforts — publications, presentations at meetings, consultations with state and local educators—will be stepped up in order to make utilizations more feasible. Special projects such as workshops for state and local assessment specialists, establishment of experimental and pilot or demonstration schools or school districts to apply the National Assessment model and materials, are in the planning stage.

Use for Educational Improvement

In contrast to the above, utilizations of the data reported through National Assessment are envisioned as taking a different form. These kinds of utilizations require data interpretation and their implications for education. In contrast to utilization which generates information, this utilization takes already generated information and uses it for educational improvement and decision-making. With the projected completion of baseline reporting in science, citizenship and writing, the entire sets of National Assessment data on achievement in three subject areas will be available for the first time. Data in other subject areas will be available as each successive year of the project is completed.

NAEP staff has traditionally considered this area as being outside the scope of its mission. However, NAEP staff can serve as a catalytic influence to promote interpretive activities, and it is this kind of an active role which it will assume. In line with its mission, a number of activities are contemplated, such as sponsoring work groups, reaction panels or study groups made up of educational specialists and lay people to study and interpret results and suggest implications; working with curriculum projects and publishers of instructional materials to make modifications and reemphases suggested by the data; encouraging professional education organizations to assume an active role in interpreting and implementing National Assessment results and other activities of a comparable nature.

Also in line with this aspect of the applications program will be the redesigning of NAEP reports to facilitate their use and interpretation as well as active dissemination through publication, conferences, etc. of known and potential applications.

National Assessment Utilizations, still in its formative stages, has generated considerable excitement about the possibilities of improving American education. The participation of educational decision-makers, curriculum specialists, instructional designers and concerned laymen in this important effort is actively solicited.

Trends in State Assessment Activity

Trends and Issues as identified in State Educational Assessment Programs, a survey conducted by the Educational Testing Service.

The comprehensive description of state assessment activity is found in a 1971 publication of the Educational Testing Service (ETS) entitled *State Educational Assessment Programs.* The overall impression that one gets from the survey is that state assessment plans and programs are currently in a highly fluid state, with new developments occurring daily. While about half the states have actual testing programs, many others are involved in various types of evaluation and assessment activities. On the basis of the data gathered, ETS has identified several trends and issues that characterize the national picture. Among the trends are the following:

1. Every state has developed or is developing a *needs assessment.* Such an assessment is necessary before a state can receive Title III, Elementary and Secondary Education Act (ESEA) funds. In many states some measurement of pupil performance is included.

2. Formulating statewide educational goals is still another task in which many of the states are engaged. A variety of methods is being used to determine statewide goals.

3. There is a notable trend in many states to apply to the management of the educational enterprise the principles of cost benefit analysis embodied in some form of planning - programming - budgeting system (PPBS) and to incorporate statewide educational assessment into such a system.

4. Statewide testing is becoming an increasingly important element of assessment.

5. Although most testing involves evaluation of basic skills (cognitive development) there is an obvious desire to broaden evaluation to include non-cognitive elements of education.

6. A significant advance in mounting state testing programs, is the commitment on the part of a number of states to assessing the outcomes of education only after accounting for the effects of community and home environment, of teachers and school programs, and of school facilities and financial resources.

As states get into assessment development a number of issues and problems arise. These are as follows:

In assessing pupil performance, a state must decide whether to use a sampling approach or an "every-pupil" approach. Settling the issue often appears to depend on how the purpose of statewide assessment is locally perceived.

Assessment planning raises the question of the balance of control between the state and the local school district.

In some states, assessment planning may be going on in administrative departments, legislative committees, and the governor's office without communication or coordination.

A question confronting some states is: How shall results relate to financial incentives?

The "handling of sensitive data," considered desirable in interpreting results, raises questions of the invasion of privacy, and accuracy of data.

In the establishment of goals there appears to be a considerable amount of confusion between the ends and means of education between process and product, between inputs and outputs, and between pupil performance objectives, staff objectives, and system objectives.

ECS Support of Program

The Steering Committee of the Education Commission of the States at its December 1971 meeting approved the use of National Assessment funds to provide a more vigorous service to help states in developing their own state assessment programs. A Department of Utilization within National Assessment was created and will be available to do the following:

1. To meet with state representatives to explain the National Assessment model—its objectives, how it operates, etc.
2. To explain how the National Assessment model might be adapted to state or local use.
3. To provide basic cost estimates to states to perform various tasks, e.g. population sampling, exercise administration, etc.
4. To assist states in utilizing NAEP assessment materials.
5. To identify consultant services to states which wish to implement NAEP technology.
6. To provide workshops and seminars for state department representatives and others.
7. To show how states can "tie-in" to National Assessment as a part of their own program.

ECS, in taking this action, is responding to many requests from states. Up to this time neither manpower nor funds were available to meet all these requests. It should be emphasized that ECS is not promoting the NAEP model or urging states to adopt it. It does, however, encourage states to plan assessment programs.

States will have varying needs and can use varying technologies of which NAEP is but one. Two organizations which have publicly announced their availability and from which states can receive specific help in planning assessment designs, customized to state needs, are the Educational Testing Service which has recently set up a State Assessment Center and the Consortium composed of Research Triangle Institute, Measurement Research Center, and American Institutes for Research.

Announcing:

ECS 6th Annual Meeting
May 17-19, 1972
Century Plaza Hotel
Los Angeles, California

ECS STEERING COMMITTEE
1971-1972

Chairman
Governor Robert W. Scott, North Carolina
Vice Chairman
Dr. Warren Hill, Ex. Dir., Connecticut Comm. for Higher Education
Chairman-Elect
Governor Winfield Dunn, Tennessee
Treasurer
Senator Bennett Katz, Maine
Members
Governor Russell W. Peterson, Delaware
Governor Walter Peterson, New Hampshire
Governor Bruce King, New Mexico
Governor Tom McCall, Oregon
Governor Luis Ferre', Puerto Rico
Governor Stanley Hathaway, Wyoming
Senator D. Robert Graham, Florida
Representative Charles Clabaugh, Illinois
Representative Darvin Allen, Kentucky
Senator Oakley Collins, Ohio
Senator Bryce Baggett, Oklahoma
Representative Max Homer, Pennsylvania
Representative Manny Brown, Wisconsin
James Stratten, Division of Apprenticeship Standards, California
Frederick P. Thieme, President, University of Colorado
Richard Kosaki, Chancellor, New Campus, University of Hawaii
Wilson Elkins, President, University of Maryland
Edward Moore, Chancellor of Higher Education, Massachusetts
Karl Grittner, High School Principal, Minnesota
Everett Keith, Executive Secretary, Missouri Teachers Association
Ewald Nyquist, State Commissioner of Education, New York
Wm. P. Robinson, Jr., Assoc. Comm. of Education, Rhode Island
Cyril Busbee, State School Superintendent, South Carolina
Darld J. Long, Executive Director, Utah School Boards Association
Robert Babcock, Provost, Vermont State College
Robert Williams, Ex. Secretary, Virginia Education Association
Mrs. Eldra Shulterbrandt, Trustee, College of the Virgin Islands
Advisory Members
Governor William Egan, Alaska
Governor James Exon, Nebraska
Governor Arch Moore, West Virginia
Representative Peter Turnham, Alabama
Representative B. G. Hendrix, Arkansas
Representative Ernest Allen, Idaho
John Loughlin, Superintendent of Public Instruction, Indiana
William J. Dodd, State Superintendent of Education, Louisiana
Rev. John Bloh, Supt. of the Archdiocese, Trenton, New Jersey
Mrs. Ray Miller, School Board Member, Fargo, North Dakota
Abner McCall, President, Baylor University, Texas
Mrs. Jerome Freiberg, Seattle, Washington

ECS SPRING STEERING COMMITTEE MEETING
MARCH 9-10, 1972
The Broadmoor, Colorado Springs, Colorado

Topics will include making state educational councils effective, state implications of the President's Commission on School Finance report and assistance to states in relation to education and property tax issues.

Reservation information can be obtained from ECS, Suite 300, Lincoln Tower, 1860 Lincoln, Denver, Colorado 80203.

Observers welcome!